View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Bath Research Portal

Citation for published version:
Kover, PX & Mott, R 2012, 'Mapping the genetic basis of ecologically and evolutionarily relevant traits in Arabidopsis thaliana', *Current Opinion in Plant Biology*, vol. 15, no. 2, pp. 212-217. https://doi.org/10.1016/j.pbi.2012.02.002

DOI:
10.1016/j.pbi.2012.02.002

Publication date:
2012

Document Version
Peer reviewed version

Link to publication

NOTICE: this is the author's version of a work that was accepted for publication in Current Opinion in Plant Biology. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Current Opinion in Plant Biology, vol 15, issue 2, 2012, DOI 10.1016/j.pbi.2012.02.002

University of Bath

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 23. Jan. 2020

Mapping the genetic basis of ecologically and evolutionarily relevant traits in *Arabidopsis thaliana*

Paula X. Kover[a*] and Richard Mott[b]

a - University of Bath, Dept of Biology and Biochemistry

Claverton Down, BA2 7AY - UK

b- Wellcome Trust Centre for Human Genetics

Roosevelt Drive, Oxford OX3 7BN - UK

* Corresponding author:

Paula X. Kover

Email: p.x.kover@bath.ac.uk

Tel:+44 (0)122 5385059

Fax: +44 (0)122 5386779

ABSTRACT

There has been a long standing interest in the relationship between genetic and phenotypic variation in natural populations, in order to understand the genetic basis of adaptation and to discover natural alleles to improve crops. Here we review recent developments in mapping approaches that have significantly improved our ability to identify causal polymorphism explaining natural variation in ecological and evolutionarily relevant traits. However, challenges in interpreting these discoveries remain. In particular, we need more detailed transcriptomic, epigenomic, and gene network data to help understand the mechanisms behind identified associations. Also, more studies need to be performed under field conditions or using experimental evolution to determine whether polymorphisms identified in the lab are relevant for adaptation and improvement under natural conditions.

INTRODUCTION

Plants, perhaps due to their limited mobility, show an impressive array of intraspecific and interspecific variation for traits such as drought, disease resistance, tolerance to salinity, etc. [1-3]. Much of the phenotypic variance seem to reflect specific adaptations to climate and seasonality and are thought to have been shaped by natural selection. Underlying all this phenotypic variation is, of course, extensive genomic, transcriptomic, and epigenomic variation. There has been a long standing interest in tapping onto this variation to improve crops and to dissect the genetic basis of adaptation.

Recent technological developments have significantly improved our ability to catalogue all this genomic diversity [e.g. 4-8]. The hope is that a complete catalogue will lead to the unravelling of the relationship between natural genetic and phenotypic variation, what is also known as the “genotype to phenotype map” [9] (Fig 1). Such a map would facilitate the identification of useful candidate genes to improve crops, allowing the identification of gene transfer from model species into crops or the identification of useful orthologs [10-11]. It will also allow a better understanding of the value of different kinds of genetic diversity for the maintenance of evolutionary potential and therefore for species’ persistence [12-13]. These are particularly pressing issues given the challenges plants (and therefore, us) may encounter in the near future when significant changes in the environment are expected [14]. Here, we will review the current approaches to identify the genetic basis of ecological and evolutionary traits and discuss whether we are getting closer to understanding them.

Mapping using experimental populations

Most traits of evolutionary and economical relevance (e.g. germination rate, competitiveness, glucosinolate content, fitness, etc) are complex, that is, they are determined by multiple loci, which may interact with each other, and they are often affected by environmental and parental effects. Thus, identifying the genetic basis of these traits has been quite challenging. Currently, the main approach to identify loci underlying natural variation in complex traits is to associate genotypic and phenotypic variation using either recombinant inbred lines (RILs) or a collection of natural accessions.

Until very recently, RILs were mainly produced by crossing two inbred accessions (genomes independently collected from natural populations) to obtain a segregating F2 population (Fig. 2a). The recombined genotypes are then inbred for 5 or more generations, to produce homozygous genotypes (RILS) that only need to be genotyped once, and later phenotyped for multiple traits under different environments [e.g. 15-16]. Marker association between mapped genotyped variants and phenotype indicate the locations of genetic factors that significantly affect the trait. However, these lines typically do not offer very good accuracy, with QTL being located to chromosomal regions varying between 5 and 50 cM (which is equivalent to 1.2 to 12 Mb in *A. thaliana*, or 9 to 90 Mb in Maize) [17]. To reduce the interval containing the QTL, more recombinations are needed. In Advanced Recombinant inbred lines (AILs, Fig 2b), two or more generations of recombinations are performed before the inbreeding starts, reducing the confidence interval for the QTL localization [18-19].

Recently, multiparent RILs have been developed: The Arabidopsis Multiparent RILs (AMPRIL) [20], and the Multiparent Advanced Genetic InterCross (MAGIC) lines [21]. The MAGIC lines were derived from intercrossing 19 natural accessions of *A. thaliana* for multiple generations (Fig. 2c). Thus, in addition to extra recombination steps, there are also a larger number of alleles segregating in the lines; as a result QTL can be mapped to significantly smaller regions (< 1 Mb). This resource has been further enhanced by the sequencing of the parental genomes [5], which revealed 3.3 million variable sites that are segregating in the MAGIC lines. The level of nucleotide diversity present among the MAGIC lines is similar to the diversity observed among 96 accessions previously genotyped [5, 8]. Also, they contain 35% of all the non-private SNPs uncovered so far by the 1001 genome project (which includes at the moment 452 accessions), and 68% of the SNPs with frequency higher than 0.05 (Cao, pers. comm.). This suggests that although only 19 accessions were used to make the MAGIC RILs, they do capture a significant proportion of the common molecular variation present in the species. While MAGIC and AMPRIL lines

allow higher precision than a traditional RIL, they may detect fewer QTL [20]. However, more data is needed to determine if this is indeed the case since only a couple of studies have yet been published, and what would be the cause of such reduction.

Mapping using natural populations

QTL maps are not the only method available to dissect complex traits; an alternative is to perform GWAS by looking for associations between SNPs and phenotypic variation across a large panel of naturally-occurring inbred accessions [4]. This has been possible due to the reduction in sequencing cost, and the development of new technologies [reviewed in 23] , leading to the genome sequencing of >400 natural accessions of *A. thaliana* ([5,22,24] and http://1001genomes.org/). The reduction in price and build up of species being sequenced will probably allow GWAS to also be pursued soon in many other species.

The advantages of using GWAS in *A. thaliana* is that in a large worldwide population, many sequence variants have arisen by mutation and many recombinations have accumulated, resulting in a situation where the linkage disequilibrium between common variants decays on average to background levels within a few kilobases [25]. Thus, a SNP association with a complex trait is likely to be very close to the causal variant. This was clearly shown by the identification of a natural polymorphism in *HKT1*as the cause of natural variation in salt accumulation, and its association with coastal populations [26]. However, there is extensive population structure in *A. thaliana*, and a large number of rare variants (frequency < 5%) [27]. Population structure means that distant variants (even on different chromosomes) are sometimes in disequilibrium with each other, causing false positive genetic associations (estimated to be about 40% in one study [28]). This can be ameliorated by control for population structure, but this can also reduce the power to detect associations [4]. In addition, rare variants affecting a phenotype are hard to detect or map accurately, first because they are unlikely to account for much of the phenotypic variation, and second, if they are of recent origin, they are likely to be in linkage disequilibrium with a larger region of the surrounding genome. GWAS has also been successfully implemented in rice and maize [29-31], illustrating how natural variation is helping identify the genetic basis of complex agronomically valuable traits.

Combining approaches

QTL analysis and GWAS have complementary characteristics: RILs control for allele frequency (the minor allele frequency should never be much under 1/ [# of parental founders]) and population structure is broken up. However, RILs experience less recombination, being limited to those accumulating in meioses during their production. Thus mapping resolution is lower, but there are fewer false positives. It has been suggested that a possible powerful approach is to use RILs and natural accessions to map the same trait simultaneously [28]. The GWAS would provide good resolution for the causal polymorphisms, and the QTL analysis would allow detection of false-positives. It is possible that some of the associations observed with GWAS and not with QTL analysis are true, but there is not sufficient power to detect it on the QTL analysis. This would result in a slightly elevated false negative rate, but that seems less of a problem when the goal is to identify some or the major loci contributing to the variation of a given loci. Another issue is the choice of RILs, since different associations may be due to alleles from different accessions. Synthetic populations derived from multiple accessions, such as the MAGIC or the AMPRIL lines [18] are likely to be particularly useful for this purpose.

Another approach pioneered in maize [32] to combine GWAS and QTL mapping, and particularly useful for species with larger genomes, is to sequence a few accession (25 in the case of maize) and produce a set of RILs between these founders. This approach, also known as nested association mapping (NAM) has led to the successful identification of loci associated with many traits, including disease resistance [31].

Further searching;Trancriptomics and Epigenetics

Despite the impressive depth of knowledge about genomic variation available in *A. thaliana*, recent studies suggest that we have only scratched the surface. Detecting associations between SNPs and phenotype is only the first step in elucidating the mechanism through which genomic variation affects phenotype. More data on quantitative and qualitative variation at the protein, RNA, and chromatin level would provide valuable additional information to unravel the genotype to phenotype map (Fig 1).

RNA-seq data from seedlings was obtained for the 19 MAGIC parental accessions, and used to re-annotate coding genes in each accession [5]. Surprisingly, it was found that a simple transfer of gene annotations from the reference Col-0 to each accession predicts about one third of the genes to be significantly disrupted in at least one accession (e.g. by a premature stop codon). However, if each accession is annotated *de novo*, backed-up by RNA seq data, it is observed that often the splicing pattern of an affected gene changed in such a way that the new transcript is still functional. Thus, using the default reference gene

annotations to predict the effects of sequence variation is fraught, and the only sure way to correct the annotations is by obtaining transcript data. This study also showed that ~9000 transcripts vary among accessions and may possibly also vary between tissues. Thus, any catalog of genes expressed in one accession of *A. thaliana* represent a biased sample, and better interpretation of associations between SNPs and phenotype will require more RNA seq across accessions and multiple tissues.

Epigenetic studies, which focus on DNA and chromatin modifications, can also provide a useful complement by providing an extra source of heritable variation that may explain some of the phenotypic variation that tend to remain unexplained in many studies of complex traits [33-35]. Intraspecific variation in epigenetic markers has been observed in a number of species [36-38], and it has been shown to affect gene expression in *A. thaliana* [39]. The importance of epigenetic marks on phenotypic expression has been demonstrated by the significant difference in plant heights among *A. thaliana* lines with identical genomes but different methylation patterns [40]. Recent advances in genomic technologies have allowed genome-wide methylation marks to be assayed efficiently [41], opening the doors for the performance of Epigenome-Wide Association Studies (EWAS). EWAS searches for association between DNA methylation marks and phenotype similarly to GWAS. EWAS has been used successfully to identify associations between methylated sites and autism in humans [42], and there are ongoing studies exploring its use in Arabidopsis.

Which genes are being identified?

It is unclear whether most loci identified so far, using the methods discussed above, are relevant to address the genetic basis of adaptation or to improve crops facing significant environmental changes. The first concern comes from the possible effect of environment in determining significant associations [43-44]. A recent study that compared associations found under laboratory and field conditions in *A. thaliana* found a very different set of loci associated flowering time (an important trait to obtain optimal fitness under varying environments) [45]. Of the 25 loci were found to be associated in the field, only 2 were previously identified to be associated with flowering time under lab conditions. A second concern is that selection experiments have suggested that genes previously identified to affect flowering time, do not mediate most of the phenotypic response to selection [21,46-47]. Such discrepancy might be due to differences in environment between QTL and selection studies; but it might reflect the existence of genetic constraints, such that the genes that respond to selection differ from those with detectable associations at a single point in time. Consequently, studies to determine the genetic basis of the response to selection, in which populations that have been allowed to evolve are resequenced, may improve our

understanding of the relationship between existing natural variation and the response to selection.

Conclusions

Statistical developments, sequencing technologies and new mapping methods have significantly improved our ability to detect natural polymorphisms that underlie ecological and evolutionary relevant traits. However, many more studies are still need to understand the adaptive process, and the mechanisms through which causal variants affect phenotype. The biggest challenge ahead is to expand the tools developed in *A. thaliana* to many other non-model species, and to use experimental evolution to understand the relationship between standing variation and evolutionary response. The development of faster and cheaper technologies to essay genomic variation at multiple levels is quickly transforming our knowledge about patterns of natural genetic variation within [48-49] and across related species [50-51]; and it will certainly transform our understanding of the evolutionary process.

Acknowledgements

We thank Y.Van der Peer and J.C. Pires for their invitation to contribute to this issue. We also acknowledge funding support from NERC and BBSRC.

Literature cited

* Of special interest

** Of outstanding interest

1. Nuismer SL, Gandon S: **Moving beyond common-garden and transplant designs: Insight into the causes of local adaptation in species interactions**. *The American Naturalist* 2008, **171**:658-668.
2. Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M: **What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation?** *The Plant Cell Online* 2009, **21**:1877-1896.
3. Koepke D, Kolb T, Adams H: **Variation in woody plant mortality and dieback from severe drought among soils, plant groups, and species within a northern Arizona ecotone**. *Oecologia* 2010, **163**:1079-1090.

4. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, et al.: **Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines**. *Nature* 2010, **465**:627-631.
* This paper presents an amazing tour de force, and firmly establis GWAS in *A. thaliana*
5. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, et al.: **Multiple reference genomes and transcriptomes for Arabidopsis thaliana**. *Nature* 2011, **477**:419-423.
** This reports the first joint study of genome-wide natural variation in genome and transcriptome sequence, and provide some very compelling arguments for the importance of annotating accessions de novo.

6. Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, et al.: **Epigenetic natural variation in *Arabidopsis thaliana***. *PLoS Biol* 2007, **5**:e174.
7. Schneeberger K, Weigel D: **Fast-forward genetics enabled by new sequencing technologies**. *Trends in Plant Science* 2011, **16**:282-288.
8. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, et al.: **Common Sequence Polymorphisms Shaping Genetic Diversity in Arabidopsis thaliana**. *Science* 2007, **317**:338-342.
9. Rockman MV: **Reverse engineering the genotype-phenotype map with natural genetic variation**. *Nature* 2008, **456**:738-744.
10. North H, Baud S, Debeaujon I, Dubos C, Dubreucq B, Grappin P, Jullien M, Lepiniec L, Marion-Poll A, Miquel M, et al.: **Arabidopsis seed secrets unravelled after a decade of genetic and omics-driven research**. *The Plant Journal* 2010, **61**:971-981.
11. Spannagl M, Mayer K, Durner J, Haberer G, Fröhlich A: **Exploring the genomes: From Arabidopsis to crops**. *Journal of Plant Physiology* 2011, **168**:3-8.
12. Hoffmann AA, Willi Y: **Detecting genetic responses to environmental change**. *Nat Rev Genet* 2008, **9**:421-432.
13. Chevin L-M, Lande R, Mace GM: **Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory**. *PLoS Biol* 2010, **8**:e1000357.
14. IPCC: **Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K and Johnson CA (eds.)**. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881pp* 2001.
15. Lister C, Dean C: **Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana***. *Plant Journal* 1993, **4**:745-750.
16. Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F: **Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis**. *Theor Appl Genet* 2002, **104**:1173-1184.
17. Alonso-Blanco C, Mendez-Vigo B, Koornneef M: **From phenotypic to molecular polymorphism involved in naturally occuring variation of plant development**. *Int. J Dev. Biol.* 2005, **49**:717-732.
18. Balint-Kurti PJ, Yang J, Esbroeck GV, Jung J, Smith ME: **Use of a maize Advanced Intercross Line for mapping of QTL for northern leaf blight resistance and multiple disease resistance.** *Crop Science* 2010, **50**:458-466.
19. Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO, et al.: **QTL Mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines**. *PLoS ONE* 2009, **4**:e4318.

20. Huang X, Paulo M-J, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA: **Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population**. *Proceedings of the National Academy of Sciences* 2011, **108**:4488-4493.
21. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich I, Purugganan M, Durrant C, Mott R: **A multiparent advanced generation intercross to fine-map quantitative traits in *Arabidopsis thaliana***. *PLoS genetics* 2009, **5**:e1000551.
** This paper presents the first multiparent set of RILs in plants.

22. Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, et al.: **Whole-genome sequencing of multiple Arabidopsis thaliana populations**. *Nat Genet* 2011, **43**:956-963.
** This study reports on the genomic variation on the impressive number of 80 natural accessions of *A. thaliana,* the largest so far.

23. Mardis E: **Next-generation sequencing methods**. *Annu. Rev. Genomics Hum. Genet* 2008, **9**:387-402.
24. Weigel D, Mott R: **The 1001 Genomes Project for Arabidopsis thaliana**. *Genome Biology* 2009, **10**:107.
25. Kim S, Plagnol V, Hu T, Toomajian C, Clark R, Ossowski S, Ecker J, Weigel D, Nordborg M: **Recombination and linkage disequilibrium in *Arabidopsis thaliana***. *Nature Genetics* 2007, **39**:1151-1155.
26. Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M, et al.: **A Coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1**. *PLoS Genet* 2010, **6**:e1001193.
27. Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Ågren J, Bossdorf O, Byers D, Donohue K, et al.: **The scale of population structure in *Arabidopsis thaliana***. *PLoS Genet* 2010, **6**:e1000843.
*Really nice work showing that although self-fertilization is high in A. thaliana, the small outcrossing rate is sufficient to introduce considerable genetic recombination after just a few generations, and there is no reason to think about this species as an unusual, exclusively selfing species.

28. Bergelson J, Roux F: **Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana**. *Nat Rev Genet* 2010, **11**:867-879.
29. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, et al.: **Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm**. *Nat Genet* 2012, **44**:32-39.
30. Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, et al.: **Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa**. *Nat Commun* 2011, **2**:467.
31. Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, et al.: **Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population**. *Nat Genet* 2011, **43**:163-168.
32. Yu J, Holland JB, McMullen MD, Buckler ES: **Genetic Design and Statistical Power of Nested Association Mapping in Maize**. *Genetics* 2008, **178**:539-551.
33. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al.: **Finding the missing heritability of complex diseases**. *Nature* 2009, **461**:747-753.

34. Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P, et al.: **Assessing the impact of transgenerational epigenetic variation on complex traits**. *PLoS Genet* 2009, **5**:e1000530.
35. Roux F, Colomé-Tatché M, Edelist C, Wardenaar R, Guerche P, Hospital F, Colot V, Jansen RC, Johannes F: **Genome-Wide epigenetic perturbation jump-starts patterns of heritable variation found in nature**. *Genetics* 2011, **188**:1015-1017.
*First study to show the importance of epigenetic variation under field conditions.

36. Herrera CM, Bazaga P: **Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis**. *New Phytologist* 2010, **187**:867-876.
37. Li Y, Shan X, Liu X, Hu L, Guo W, Liu B: **Utility of the methylation-sensitive amplified polymorphism (MSAP) marker for detection of DNA methylation polymorphism and epigenetic population structure in a wild barley species (<i>Hordeum brevisubulatum</i>)**. *Ecological Research* 2008, **23**:927-930.
38. Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, et al.: **Epigenetic natural variation in *Arabidopsis thaliana***. *PLoS Biol* 2007, **5**:e174.
39. Zhang X, Shiu S, Cal A, Borevitz JO: **Global analysis of genetic, epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana* using whole genome tiling arrays**. *PLoS Genet* 2008, **4**:e1000032.
40. Johannes F, Colomé-Tatché M: **Quantitative epigenetics through epigenomic perturbation of isogenic lines**. *Genetics* 2011, **188**:215-227.
41. Henderson IR, Chan SR, Cao X, Johnson L, Jacobsen SE: **Accurate sodium bisulfite sequencing in plants**. *Epigenetics* 2010, **5**:47-49.
42. Rakyan VK, Down TA, Balding DJ, Beck S: **Epigenome-wide association studies for common human diseases**. *Nat Rev Genet* 2011, **12**:529-541.
43. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, et al.: **The genetic architecture of maize flowering time**. *Science* 2009, **325**:714-718.
**Impressive work including ~ 1 million plants, showed very nicely that what looked like a traits with a few QTL of large effect, become a trait with many loci of smaller effect when there is significant power.

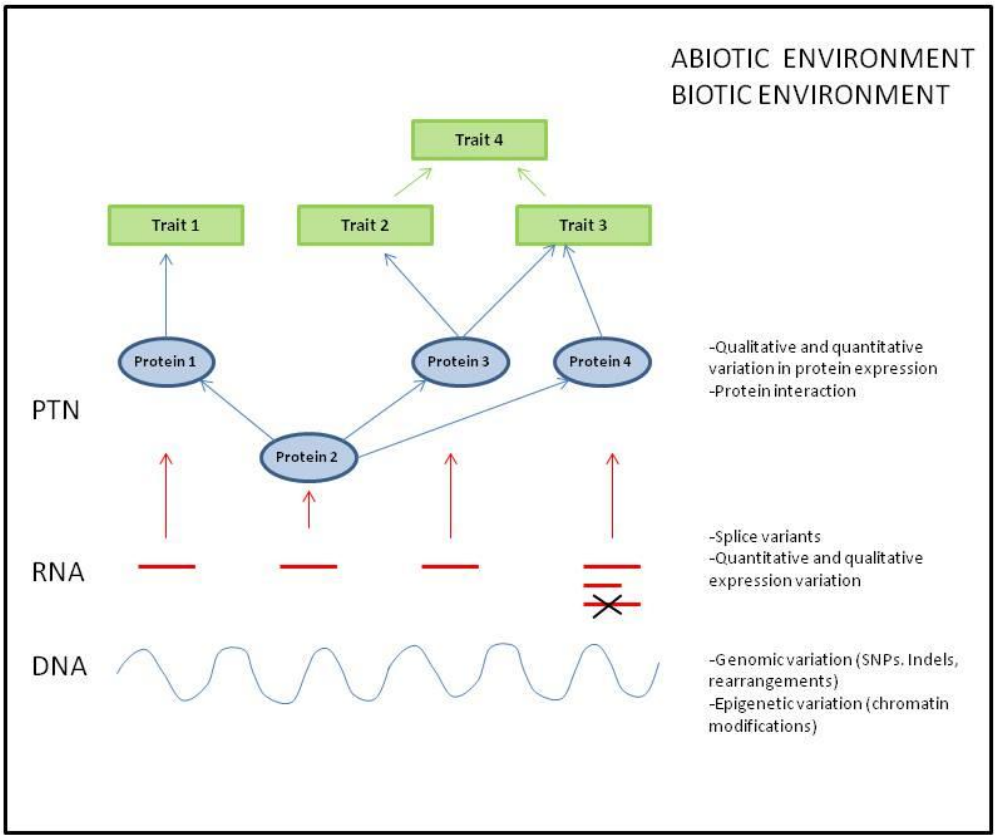
44. Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM: **A map of local adaptation in *Arabidopsis thaliana***. *Science* 2011, **334**:86-89.
45. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F: **Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature**. *PLoS Genet* 2010, **6**:e1000940.
46. Scarcelli N, Kover PX: **Standing genetic variation in *FRIGIDA* mediates experimental evolution of flowering time in Arabidopsis**. *Molecular Ecology* 2009, **18**:2039-2049.
47. Kover PX, Rowntree JK, Scarcelli N, Savriama Y, Eldridge T, Schaal BA: **Pleiotropic effects of environment-specific adaptation in Arabidopsis thaliana**. *New Phytologist* 2009, **183**:816-825.
48. Branca A, Paape TD, Zhou P, Briskine R, Farmer AD, Mudge J, Bharti AK, Woodward JE, May GD, Gentzbittel L, et al.: **Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume Medicago truncatula**. *Proceedings of the National Academy of Sciences* 2011, **108**:E864–E870.
49. Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, Li M-W, He W, Qin N, Wang B, et al.: **Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection**. *Nat Genet* 2010, **42**:1053-1059.

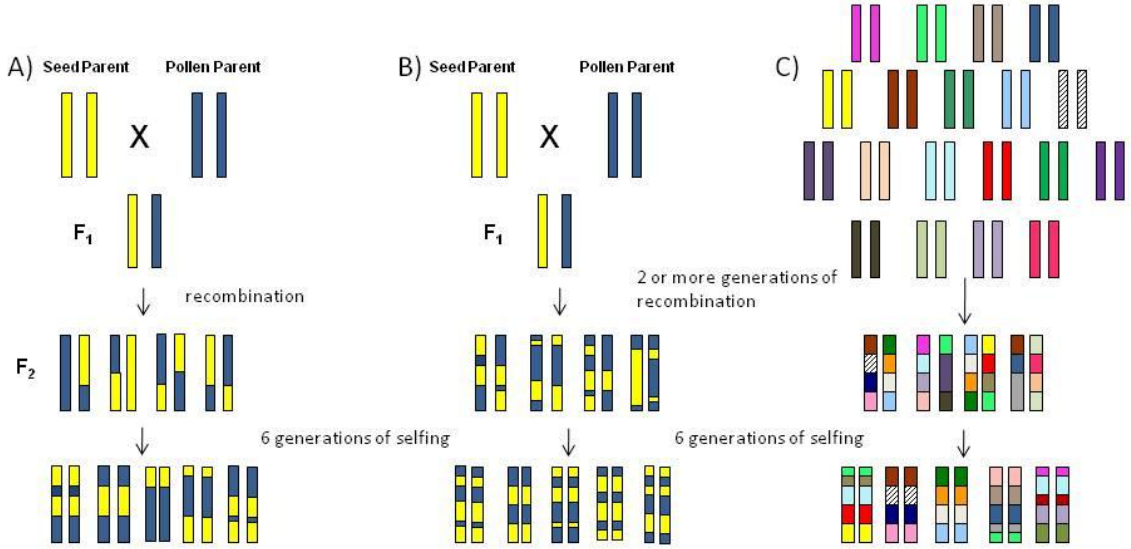
50. Dassanayake M, Oh D-H, Haas JS, Hernandez A, Hong H, Ali S, Yun D-J, Bressan RA, Zhu J-K, Bohnert HJ, et al.: **The genome of the extremophile crucifer Thellungiella parvula**. *Nat Genet* 2011, **43**:913-918.
51. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al.: **The Arabidopsis lyrata genome sequence and the basis of rapid genome size change**. *Nat Genet* 2011, **43**:476-481.

Figure legends

Fig 1. Schematic illustration of the main sources of heritable variation influencing the genotype to phenotype map. Trait 1 illustrates a simple qualitative trait, such as disease resistance in plants where the expression of a receptor protein may be sufficient to elicit defense responses against biotic components of the environment. In contrast, trait 2 and 4 illustrate complex traits mediated by more than on gene. Trait 4 is an example of a composite trait such as fitness, that depends on the combination of multiple traits such as fruit number and number of seeds per fruit, which can have part of its genetic basis in common (Protein 3) and some independent (Protein 4). In reality, we expect much more complex relationships.

Fig. 2. Illustration of the different ways in which Recombinant Inbred Lines (RILs) for mapping complex traits can be produced. Panel A shows a traditional RIL produced from the F2 of two accessions; panel B shows Advanced Intercross lines, and how the extra generation(s) of recombination reduces the size of the chromosome fragments; and panel C shows a Multiparent Advanced Genetic InterCross (MAGIC) lines, where there are multiple parental accessions and extra generations of recombination.


ABIOTIC ENVIRONMENT
BIOTIC ENVIRONMENT
Trait 4
Trait 1
Trait 2
Trait 3
Protein 1
Protein 3
Protein 4
Protein 2
PTN
RNA
DNA
-Qualitative and quantitative variation in protein expression
-Protein interaction
-Splice variants
-Quantitative and qualitative expression variation
-Genomic variation (SNPs. Indels, rearrangements)
-Epigenetic variation (chromatin modifications)


A)
Seed Parent
Pollen Parent
X
F1
recombination
F2
6 generations of selfing
B)
Seed Parent
Pollen Parent
X
F1
6 generations of selfing
C)
2 or more generations of recombination